Ontology Alignment for Real-World Applications*

Isabel F. Cruz William Sunna Anjli Chaudhry
Department of Computer Science
University of Illinois at Chicago
[ifc,wsunna,achaudhr]@cs.uic.edu

1 Introduction

An ontology-driven approach to data integration relies on the alignment of the concepts of a global ontology that describe the domain, with the concepts of the ontologies that describe the data in the local databases. Once the alignment between the global ontology and each of the local ontologies is established, users can potentially query hundreds of databases using a single query that hides the underlying heterogeneities. Using our approach, querying can be easily extended to a new database by aligning its ontology with the global one. For this purpose, we have designed and implemented a software tool to align ontologies. The output of this tool is a set of mappings between concepts, which will be used to produce the queries to the local databases once a query is formulated on the global ontology.

2 Ontology Alignment

Our application domain is provided by the Wisconsin Land Information System (WLIS) and focuses on Land Use Data. The land use database system that we consider stores information about land parcels in XML documents (the local databases). Land use categories include *agriculture, commerce, industry, institutions*, and *residences*.

In this paper, we represent the ontologies as trees. The vertices of the trees correspond to concepts in the ontology. Throughout the examples, the left tree represents the global ontology and the right tree represents the local ontology.

*This research was supported in part by the National Science Foundation under Awards EIA-0091489 and ITR IIS-0326284.

An important step in the data integration process is ontology alignment, the identification of semantically related entities in different ontologies. Related entities are mapped to one another using different kinds of mappings: *exact* (the connected vertices are semantically equivalent), *approximate* (the connected vertices are semantically approximate), *null* (the vertex in the global ontology does not have a semantically related vertex in the local ontology), *superset* (the vertex in the global ontology is semantically a superset of the vertex in the local ontology), and *subset* (the vertex in the global ontology is semantically a subset of the vertex in the local ontology).

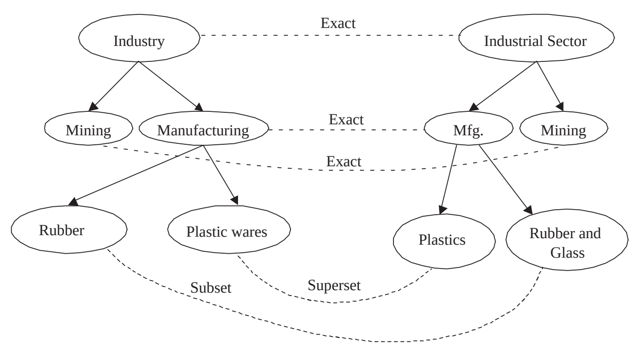

Figure 1: Mapping types.

Figure 1 illustrates several mappings between vertices in two ontologies for land use patterns. The vertices corresponding to *Industry, Mining* and *Manufacturing* in the global ontology can be mapped respectively to those corresponding to *Industrial Sector, Mining,* and *Mfg.* in the local ontology. In the global ontology, the vertex *Plastic wares* denotes entities that are made of plastic or glass. However, in the local ontology, there is a vertex *Plastics* and another vertex *Rubber and Glass*, which denotes manufactured objects made of rubber or glass. Therefore, the *Manufacturing* and the *Mfg.* vertices are aligned. Similarly, the two *Mining* vertices are also aligned. *Manufacturing* is semantically equivalent to *Mfg.*, as both denote a collection of industries producing plastics, glass, and rubber products. Hence, this mapping is of type *exact* as denoted in the mapping from the *Manufacturing* vertex to the *Mfg.* vertex. *Plastic wares* is semantically a *superset* of the *Plastics* vertex and *Rubber* is semantically a *subset* of the *Rubber and Glass* vertex.

3 Semi-automatic Alignment

To allow for the semi-automatic processing of the ontology alignment, we propose a framework that defines the values associated with the vertices of the ontology in two possible ways: as functions of the values of the children vertices or of the user input.

We need to establish two assumptions to guarantee the correctness of the deduction process. The first one is that the specialization of a vertex in the ontology must be total, that is, each higher-level concept must belong to a lower-level concept in the hierarchy. The second one is that "bowties" [2], which are inversions in the order of the two ontologies that are being aligned, do not occur.

For example, in Figure 2, vertices *b* and *c* in the global ontology are mapped using mapping types *exact* and *superset* to vertices *e* and *f* in the local ontology. The mapping type between their parents *a* and *d* can be deduced to be *superset* based on the mapping between the children. In this case, the vertices of the local ontology have corresponding vertices in the global ontology. We have devised deduction operations when this is not the case.

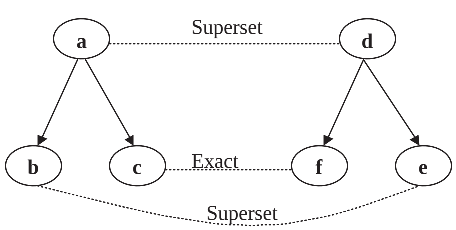

Figure 2: Deduction operation.

4 Agreement Maker

The *agreement maker* is a software tool that is used to create the mappings between the global ontology and a local ontology and generate an agreement document containing those mappings, which is used by the query processor. The query processor maps queries expressed using concepts of the global ontology to queries that use concepts of the local ontologies [1].

The local expert maps the global ontology to the local ontology with the user interface provided by the agreement maker. That user interface is shown in Figure 3. The semi-automatic alignment methodology of Section 3 has been integrated into the agreement maker tool to simplify the task of aligning large ontologies.

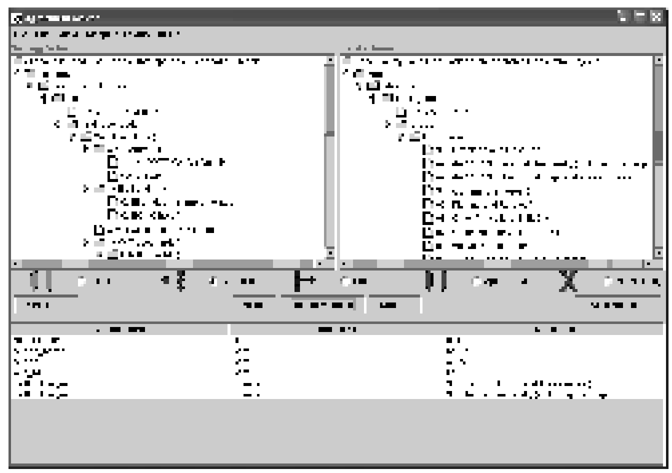

Figure 3: User interface showing mappings.

References

[1] I. F. Cruz, A. Rajendran, W. Sunna, and N. Wiegand. Handling Semantic Heterogeneities using Declarative Agreements. In *International ACM GIS Symposium*, pages 168–174, 2002.

[2] E. Hovy. Combining and Standardizing Large-Scale, Practical Ontologies for Machine Translation and Other Uses. In *First International Conference on Languages Resources and Evaluation (LREC)*, 1998.